ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

UDC 528.854.4:519.237.8 **doi: 10.32620/reks.2024.2.06**

Artem ANDREIEV[1], Leonid ARTIUSHYN[2]

[1] ***Scientific Centre for Aerospace Research of the Earth, Kyiv, Ukraine***
[2] ***State Research Institute of Aviation, Kyiv, Ukraine***

IMPROVEMENT OF LAND COVER CLASSIFICATION ACCURACY BY TRAINING SAMPLE CLUSTERING

*The **subject of this** article is land cover classification based on geospatial data. The supervised classification methods are appropriate for most of the thematic tasks of remote sensing because they provide the opportunity to set the characteristics of the initial classes in the form of a training sample set, in contrast to unsupervised methods. There are many approaches to processing such a set; however, their common disadvantage is that they do not consider the factor of training sample separability. This characteristic indicates the extent to which signatures representing different classes do not overlap. A low degree of separability is inherent in high-level training sample mixing. Thus, separability affects classification accuracy. One possible ways to increase separability is training sample clustering. Considering the above, the **goal** of this study is to develop a training sample clustering technique to improve land cover classification accuracy by increasing the separability of training samples. The **tasks** of this work are as follows: 1) develop a method for training sample separability assessment; 2) develop a training sample clustering technique based on training sample separability; 3) test the effectiveness of the developed technique by applying it to experimental land cover classification. In the experiments, two land cover classifications were obtained for each of the two selected study areas (i.e., one before and another after training sample clustering. Six land cover classes were defined for each experiment. The training samples were selected for each class. **Conclusions.** After the application of the developed technique, an increase in the separability of the training samples was evidenced by the developed separability index. In turn, this approach led to an improvement in land cover classification. For the first experiment, this was evidenced by an increase in the overall accuracy and kappa coefficient by 20% (from 63 to 83%) and 21% (from 60% to 81%), respectively. In the second experiment, the increase was 4% (from 77% to 81%) and 5% (from 66% to 71%), respectively.*

Keywords: *classification; supervised classification; unsupervised classification; clustering; remote sensing; training sample; training sample separability.*

1. Introduction

1.1. Motivation

Land cover classification has a broad range of applications in remote sensing [1]. It provides spatially explicit categorized information for environmental monitoring [2,3], land cover change detection [4], analysis of urban development [5,6], landmine detection [7], and fossil fuel exploration [8]. Moreover, land cover classification techniques play a crucial role in complex interdisciplinary problems in achieving sustainable development goals [9, 10], primarily combating climate change and its impacts [11], reversing land degradation [12], halting biodiversity loss [13], protecting water-related ecosystems for safe water supply [14], and conducting geoenvironmental hazard assessments [15].

1.2. State-of-the-art

Currently, many classification techniques have been developed, and they are mainly divided into two groups: supervised and unsupervised [16]. Supervised classification and, however, is most suitable for a large number of thematic tasks because it allows setting the characteristics of the original classes, unlike unsupervised classification. Such characteristics are provided by the training sample set.

The training sample set comprises each class sample. In turn, such a sample is presented in the form of corresponding signatures defined in each layer of the input geospatial data. Among the approaches to training sample preprocessing, the following can be distinguished: cluster sampling [17], approaches to reduce the size of the input data [18], noisy image processing [19, 20], and approaches that define mislabeled training data [21]. Along with the above approaches, we highlight the ones that aim to change the data structure, namely, image contour segmentation [22], synthesis of neural network structure [23], ranking and selection of different sampling strategies [24] and iterative clustering for training sample refinement [25].

However, most existing approaches do not consider the factor of training sample separability, which affects classification accuracy. This characteristic indicates the extent to which signatures representing different classes do not overlap. A low degree of separability is inherent in high-level training sample mixing. In turn, this leads to a significant number of misclassified objects in the classification. Thus, the training sample separability directly affects the classification accuracy [26].

1.3. Objective and Approach

The aim of the present study is to improve land cover classification accuracy. This can be achieved by increasing the separability of training samples. Thus, a method for training sample separability assessment was developed. This method is the basis of the proposed technique, which implies that training sample separability increases via training sample clustering.

Considering the above, the paper structure consists of the following sections.

In the section "Methods", we describe the method of training sample separability assessment. This method is the basis of the proposed training sample clustering technique, which is also presented in this section.

The section "Experiments" implies performing two land cover classifications before and after applying the developed technique to each of the two selected study areas.

Finally, the section "Conclusions" briefly describes the developed methods, obtained experiment results and further research aims.

2. Methods

Training sample separability assessment. Because this technique is the optimization based on training sample separability, the object function is the developed separability index of the training sample (SITS).

SITS is the result of the training sample separability assessment, and its stepwise algorithm is described in Figure 1.

The first step involves classifier training on the input training sample set. Notably, the supervised classification method must be the same as that selected for further land-cover classification. This is because the separability depends not only on the training sample structure but also on the supervised classification method.

In the second step, each signature in the training sample set is classified by the obtained classifier.

The third step is the formation of a confusion matrix [27] for the classification obtained in the previous step.

The fourth and final step is the SITS calculation. This index can be calculated for the entire training sample and two separate classes.

The SITS of the two classes ($SITS_{pair}$) is the average arithmetic value of the sensitivity and specificity indicators [28]. Sensitivity was calculated using the following formula:

$$\text{sensitivity} = \frac{x_{ii}}{x_{ii} + x_{ji}},$$

where x_{ji} is the number of class j signatures classified as class i.

The following formula corresponds to the specificity calculation:

$$\text{specificity} = \frac{x_{jj}}{x_{jj} + x_{ji}}.$$

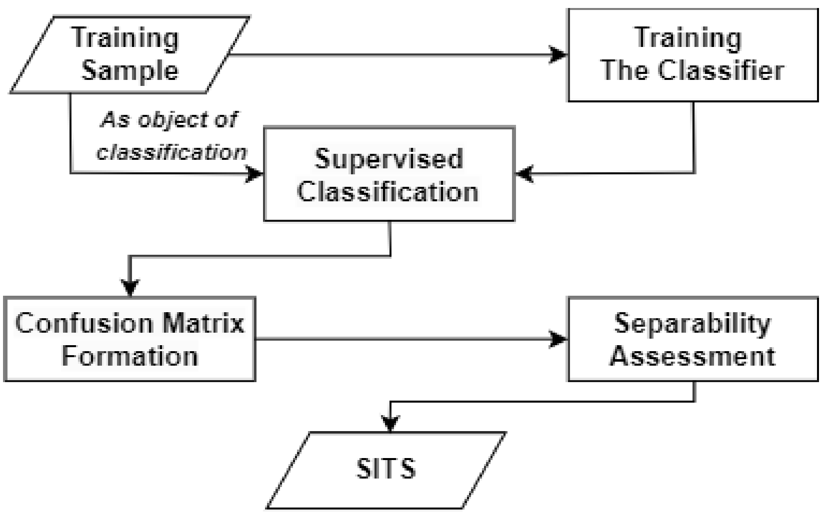

Fig. 1. The algorithm of the training sample separability assessment

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

Thus, the $SITS_{pair}$ calculation formula has the following formula:

$$SITS_{pair} = \frac{sensitivity + specificity}{2} . \tag{1}$$

SITS of the entire training sample set ($SITS_{overall}$) was calculated using the following formula:

$$SITS_{overall} = \frac{\sum_{i=1}^{K} x_{ii}}{\sum_{i=1}^{K}\sum_{j=1}^{K} x_{ij}} , \tag{2}$$

where K is the number of classes.

This index quantifies the separability of the training sample set by measuring the ratio of correctly classified signatures to the total number of signatures. In other words, the $SITS_{overall}$ equals the overall accuracy (OA) [27] based on the confusion matrix obtained in the previous step.

The values of the considered indices range from 0 to 1. At the same time, value 0 indicates that the training sample is entirely mixed (minimum separability), and value 1 corresponds to the training sample, which is entirely separable (maximum separability).

Training sample clustering technique. The developed technique assumes that only centroid methods of unsupervised classification (i.e. K-means, K-medians, along with others) [29] are considered. Thus, the goal of the proposed technique can be defined as finding the optimal number of clusters for each class of the training sample. The optimal number of clusters is one that provides the corresponding clustered set of the training sample with the highest value of the $SITS_{overall}$ among all other options. This index is calculated using equation (2). The algorithm of the developed technique is illustrated in a flowchart (Figure 2).

This algorithm is an iterative procedure. In turn, each iteration contains two steps.

The first step of the iteration is to calculate $SITS_{pair}$ for each pair of classes of the training sample using the formula (1). Then, the pair with the lowest value of this indicator is selected.

The second step involves finding the optimal number of clusters for pairs of classes selected in the previous step. For this reason, the number of clusters of these two classes increases from 1 to that number, at which the $SITS_{overall}$ value increase stops.

The iterations are then repeated without considering the pair selected for each iteration.

Such two-step iterations continue until at least one of the following stopping criteria is met:

1) if the $SITS_{overall}$ value equals 1;

2) searching for the optimal number of clusters for all consecutive pairs of classes indicates no increase in the $SITS_{overall}$.

This procedure results in the optimal structure of the training sample with the highest separability among all considered options. The obtained training sample set was used for further classification.

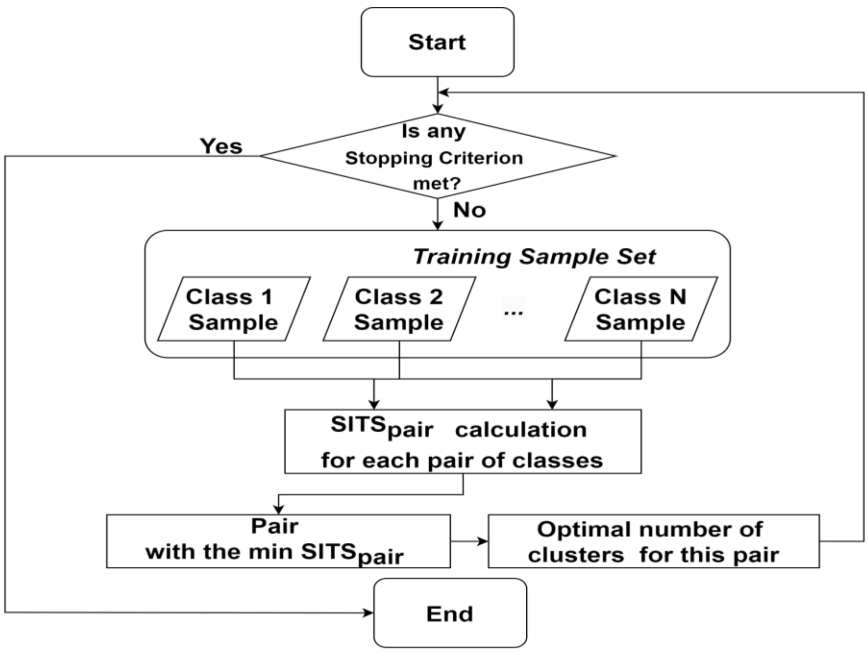

Fig. 2. The algorithm of the training the sample clustering technique

3. Experiments

Two experiments were conducted to test the effectiveness of the developed technique. Each stage consisted of performing two land cover classifications before and after applying the obtained technique. The first experiment's study area was the Ivano-Frankivsk region's test site (Figure 3, b), and the second experiment's study area was Shatsk National Natural Park (Figure 3, e).

Six broad land cover classes were defined for the first experiment: artificial surfaces, crops, grasslands, tree-covered areas, water bodies, and bare rocks. The following six classes were selected for the second experiment: artificial surfaces, crops, grasslands, tree-covered areas, water bodies, and wetlands. The training samples were selected for each of the aforementioned classes.

Input data. The data set for the first experiment included seven bands of Landsat-OLI8 satellite image (acquired on August 9, 2018) and three spectral indices (namely Normalized Difference Vegetation Index (NDVI), Normalized Difference Build-up Index (NDBI) and Build-Up Index (BUI) [30]). The second experiment data set contained ten spectral bands of the Sentinel-2 satellite image acquired on June 1, 2018.

Technique application. The maximum likelihood and Mahalanobis distance [31] were selected for the first and second experiments, respectively. In addition, K-means [29] was selected as an unsupervised classification method for training sample clustering in both experiments.

First, the initial training sample separability was assessed in each experiment. The $SITS_{overall}$ values equal 0.91 and 0.92, respectively.

Second, the training sample was clustered. The optimal number of clusters for each input class was determined for the first experiment to be 2, 5, 4, 2, 1, 2; for the second experiment to be 10, 3, 1, 4, 4, and 6. The $SITS_{overall}$ values of the obtained training samples were 0.95 and 0.99, respectively.

Finally, the corresponding classifications were performed. The initial and final classification maps for the first experiment are shown in Figures 3, a and 3, c, respectively. The classification maps for the second experiment are shown in Figures 3, d and 3, f, respectively.

Accuracy assessment. The assessment of classification accuracy involved independent verification of the initial and final land cover maps. For this purpose, the test sample set for the first experiment contained 60 pixels for each land cover map and 355 pixels for the second experiment. Satellite images (QuickBird) of high spatial resolution were used as reference data.

Metrics such as OA and kappa coefficient were selected for the classification accuracy assessment [25]. The accuracy assessment results and the SITS values are listed in Table 1.

Table 1

SITS, OA and kappa coefficients

#	SITS initial	SITS final	OA initial	OA final	Kappa initial	Kappa final
1	0.91	0.94	63	83	60	81
2	0.92	0.99	77	81	66	71

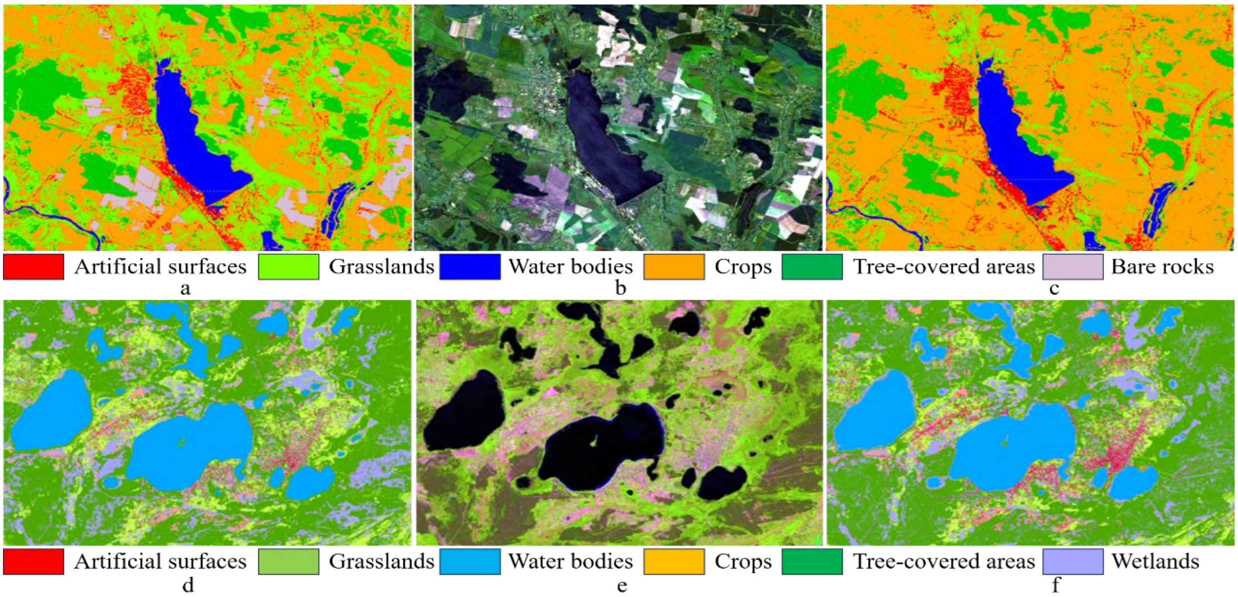

Fig. 3. First experiment: initial classification (a), Landast-OLI8 image of the study area (b), final classification (c); Second experiment: initial classification (d), Sentinel-2 image of the study area (e) and final classification (f)

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

4. Discussion

The aim of the developed optimization technique is to improve land cover classification accuracy. This goal is proposed to be achieved by increasing the separability of the training sample. Thus, the experimental result should be considered in terms of classification accuracy and training sample separability.

Two experiments confirmed the improved land cover classification accuracy after applying the proposed technique along with increased training sample separability. The SITS increased by 3% (from 91% to 94%) in the first experiment and by 7% (from 92% to 99%) in the second experiment. The OA and kappa coefficient values. For the first experiment, by 20% (from 63 to 83%) and 21% (from 60% to 81%); for the second experiment, by 4% (from 77% to 81%) and 5%(from 66% to 71%). An increase in the OA and kappa coefficients indicated an improvement in the land cover classification. For the first experiment, by 20% (from 63 to 83%) and 21% (from 60% to 81%); for the second experiment, by 4% (from 77% to 81%) and 5% (from 66% to 71%).

5. Conclusions

In this paper, we have presented a training sample clustering technique. The purpose of this technique was to increase the separability of the training sample set. Because separability directly affects classification accuracy, the technique increases the accuracy along with the separability.

In order to assess separability, an appropriate method for training sample separability assessment was developed and presented.

Further research should aim at applying the proposed technique to other classification methods.

Contributions of authors: conceptualization, methodology – **Artem Andreiev, Leonid Artiushyn;** formulation of tasks, analysis **– Leonid Artiushyn;** geospatial data processing **– Artem Andreiev;** writing – original draft preparation, writing – review and editing **– Artem Andreiev, Leonid Artiushyn.**

Conflict of Interest

The authors declare that they have no conflict of interest in relation to this research, whether financial, personal, authorship or otherwise, that could affect the research and its results presented in this paper.

Financing

This study was conducted without financial support.

Data Availability

Data will be made available upon reasonable request.

Use of Artificial Intelligence

The authors confirm that they did not use artificial intelligence technologies in their work.

All authors have read and agreed to the publication of the finale version of this manuscript.

References

1. Melesse, A. M., Weng, Q., Thenkabail, P. S., & Senay, G. B. Remote sensing sensors and applications in environmental resources mapping and modelling. *Sensors*, 2007, vol. 7, no. 12, pp. 3209-3241. DOI: 10.3390/s7123209.

2. Zaitseva, E., Stankevich, S., Kozlova, A., Piestova, I., Levashenko, V., & Rusnak, P. Assessment of the risk of disturbance impact on primeval and managed forests based on earth observation data using the example of Slovak Eastern Carpathians. *IEEE Access*, 2021, vol. 9, pp. 162847-162856. DOI: 10.1109/access.2021.3134375.

3. Kozlova, A., Stankevich, S., Svideniuk, M., & Andreiev, A. Quantitative Assessment of Forest Disturbance with C-Band SAR Data for Decision Making Support in Forest Management. *In: Lecture notes on data engineering and communications technologies*, 2021, pp 548-562. DOI: 10.1007/978-3-030-82014-5_37.

4. Stankevich, S. A., & Kozlova, A. A. Long-term land cover change computer-aided mapping by remote sensed imagery, *2015 International Conference on Information and Digital Technologies*, Zilina, Slovakia, 2015, pp. 339-341. DOI: 10.1109/DT.2015.7222994.

5. Kozlova, A., Khyzhniak, A., Piestova, I., & Andreiev, A. Synergetic use of Sentinel-1 and Sentinel-2 data for analysis of urban development and green spaces. *17th International Conference on Geoinformatics - Theoretical and Applied Aspects*, 2018, vol. 2018, pp. 1-6. DOI: 10.3997/2214-4609.201801846.

6. Eddahby, L., Kozlova, A., Popov, M., Lubskiy, N., Mezzane, D., Lukynchuk, I., Larabi, A., & Ibouh, H. Synergetic use of Sentinel-1 and Sentinel-2 data for extraction of built-up area in a rocky desert oasis, example for Draa Tafilalt, south-east of Morocco. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2019, vol. 42, no. 4, pp. 65-68. DOI: 10.5194/isprs-archives-xlii-4-w12-65-2019.

7. Popov, M. A., Stankevich, S. A., Mosov, S. P., Titarenko, O. V., Dugin, S. S., Golubov, S. I., & Andreiev, A. A. Method for Minefields Mapping by Imagery from Unmanned Aerial Vehicle. *Advances in*

Military Technology, 2022, vol. 17, no. 2, pp. 211-229. DOI: 10.3849/aimt.01722.

8. Popov, M. A., Topolnytskyi, M. V., Titarenko, O. V., Stankevich, S. A., & Andreiev, A. A. Forecasting gas and oil potential of subsoil plots via co-analysis of satellite, geological, geophysical and geochemical information by means of subjective logic. *WSEAS Transactions on Computer Research*, 2020, vol. 8, pp. 90-101. DOI: 10.37394/232018.2020.8.11.

9. Andries, A., Morse, S., Murphy, R., Lynch, J, Woolliams, E., & Fonweban, J. Translation of Earth observation data into sustainable development indicators: An analytical framework. *Sustainable Development*, 2018, vol. 27, no. 3, pp. 366-376. DOI: 10.1002/sd.1908.

10. Scott, G., & Rajabifard, A. Sustainable development and geospatial information: a strategic framework for integrating a global policy agenda into national geospatial capabilities. *Geospatial Information Science*, 2017, vol. 20, no. 2, pp. 59-76. DOI: 10.1080/10095020.2017.1325594.

11. Popov, M., Stankevich, S., Kostyuchenko, Y., & Kozlova, A. Analysis of Local Climate Variations Using Correlation between Satellite Measurements of Methane Emission and Temperature Trends within Physiographic Regions of Ukraine. *International Journal of Mathematical, Engineering and Management Sciences*, 2019, vol. 4, no. 2, pp. 276-288. DOI: 10.33889/IJMEMS.2019.4.2-023.

12. Dubovyk, O. The role of Remote Sensing in land degradation assessments: opportunities and challenges. *European Journal of Remote Sensing*, 2017, vol. 50, no. 1, pp. 601-613. DOI: 10.1080/22797254. 2017.1378926.

13. Popov, M., Kussul, N., Stankevich, S., Kozlova, A., Shelestov, A., Kravchenko, O., Korbakov, M., & Skakun, S. Web service for biodiversity estimation using remote sensing data. *International Journal of Digital Earth*, 2008, vol. 1, no. 4, pp. 367-376. DOI: 10.1080/17538940802483745.

14. Eddahby, L., Popov, M., Stankevich, S., Kozlova, A., Svideniuk, M., Mezzane, D., Lukyanchuk, I., Larabi, A., & Ibouh, H. Assessing vegetation structural changes in oasis agro-ecosystems using Sentinel-2 image time series: case study for Drâa-tafilalet region Morocco. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2019, vol. 42, no. 4, pp. 69-73. DOI: 10.5194/isprs-archives-xlii-4-w12-69-2019.

15. Stankevich, S., Piestova, I., Kozlova, A., & Titarenko, O. Satellite Radar Interferometry Processing and Elevation Change Analysis for Geoenvironmental Hazard Assessment. *Techniques for Disaster Risk Management and Mitigation*, 2020, pp. 127-139. DOI: 10.1002/9781119359203.ch10.

16. Sathya, R., & Abraham, A. Comparison of supervised and unsupervised learning algorithms for pattern classification. *International Journal of Advanced Research in Artificial Intelligence*, 2013, vol. 2, no. 2. DOI: 10.14569/ijarai.2013.020206.

17. Sedgwick, P. Cluster sampling. *BMJ*, 2014, article no.1215. DOI: 10.1136/bmj.g1215.

18. Islam, R., Siddiqa, A., Afjal, M. I., Uddin, M. P., & Ulhaq, A. Hyperspectral image classification via information theoretic dimension reduction. *Remote Sensing*, 2023, vol. 16, no. 5, article no. 807. DOI: 10.3390/rs15041147.

19. Naumenko, V., Kovalenko, B., & Lukin, V. BPG-based compression analysis of Poissonnoisy medical images. *Radioelectronic and Computer Systems*, 2023, vol. 3, pp. 91-100. DOI: 10.32620/reks.2023.3.08.

20. Rebrov, V., & Lukin, V. Post-processing of compressed noisy images by BM3D filter. *Radioelectronic and Computer Systems*, 2023, vol. 4, pp. 100-111. DOI: 10.32620/reks.2023.4.09.

21. Han, H., Li, D., Liu, W., Zhang, H., & Wang, J. High dimensional mislabeled learning. *Neurocomputing*, 2024, vol. 573, article no. 127218. DOI: 10.1016/j.neucom.2023.127218.

22. Zahorodnia, D., Pigovsky, Y., & Bykovyy, P. Canny-based method of image contour segmentation. *International Journal of Computing*, 2016, vol. 15, no. 3, pp. 200-205. DOI: 10.47839/ijc.15.3.853.

23. Marakhimov, A. R., & Khudaybergenov, K. K. Approach to the synthesis of neural network structure during classification. *International Journal of Computing*, 2020, vol. 19, no, 1, pp. 20-26. DOI: 10.47839/ijc.19.1.1689.

24. Dehkordi, A. T., Zoej, M. J. V., Ghasemi, H., Ghaderpour, E., & Hassan, Q. K. A New Clustering Method to Generate Training Samples for Supervised Monitoring of Long-Term Water Surface Dynamics Using Landsat Data through Google Earth Engine. *Sustainability*, 2022, vol. 14, no. 13, article no. 8046. DOI: 10.3390/su14138046.

25. Clark, A., Phinn, S., & Scarth, P. Pre-Processing training data improves accuracy and generalisability of convolutional neural network based landscape semantic segmentation. *Land*, 2022, vol. 12, no. 7, article no. 1268. DOI: 10.3390/land12071268.

26. Andreiev, A., Kozlova, A., Artiushyn, L., & Sedlacek, P. Enhancement of Land Cover Classification by Geospatial Data Cube Optimization. *Computer Modeling and Intelligent Systems*, 2024, vol. 3702, pp.292-304.

27. Popov, M. O. Metodologiya otsenki tochnosti klassifikatsii ob"yektov na kosmicheskikh izobrazheniyakh [Methodology of accuracy assessment of classification of objects on space images]. *Journal of Automation*

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

and Information Sciences, 2007, no. 1, pp. 97-103. Available at: https://www.researchgate.net/publication/245407157_Methodology_of_Accuracy_Assessment_of_Classification_of_Objects_on_Space_Images?enrichId=rgreq-391ade514bc4417ce4facae2236a8bcd-XXX&enrichSource=Y292ZXJQYWdlOzI0NTQwNzE1NztBUzozMDE0NDc5MzA0OTkwNzNAMTQ0ODg4MjE5NTQyMw%3D%3D&el=1_x_2&_esc=publicationCoverPdf. (accessed 12.2.2024) (In Russian)

28. Vujović, Ž. Classification model evaluation metrics. *International Journal of Advanced Computer Science and Applications/International Journal of Advanced Computer Science & Applications*, 2021, vol. 12, no. 6. DOI: 10.14569/ijacsa.2021.0120670.

29. Omran, M. G., Engelbrecht, A. P., & Salman, A. A. An overview of clustering methods. *Intelligent Data Analysis*, 2007, vol. 11, no. 6, pp. 583-605. DOI: 10.3233/ida-2007-11602.

30. Shah, S. A., Kiran, M., Nazir, A., & Ashrafani, S. H. Exploring NDVI and NDBI relationship using Landsat 8 OLI/TIRS in Khangarh Taluka, Ghotki. *Malaysian Journal of Geosciences*, 2022, vol. 6, no. 1, pp. 8-11. DOI: 10.26480/mjg.01.2022.08.11.

31. De Souza, C., Mercante, E., Prudente, V., & Justina, D. Methods of performance evaluation for the supervised classification of satellite imagery in determining land cover classes. *Cienc. Inv. Agr.*, 2013, vol. 40, no. 2, pp. 419-428. DOI: 10.4067/s0718-16202013000200016.

Received 06.02.2024, Accepted 15.04.2024

ПІДВИЩЕННЯ ТОЧНОСТІ КЛАСИФІКАЦІЇ ЗЕМНИХ ПОКРИВІВ ШЛЯХОМ КЛАСТЕРИЗАЦІЇ НАВЧАЛЬНОЇ ВИБІРКИ

А. А. Андреєв, Л. М. Артюшин

Предметом вивчення в статті є класифікація земних покривів. Як відомо, саме керовані методи класифікації є актуальними для більшості тематичних задач дистанційного зондування Землі, оскільки вони дають можливість задати характеристики вихідних класів у вигляді навчальної вибірки на відміну від некерованих методів. Існує значна кількість підходів до обробки набору навчальної вибірки, але їхнім спільним недоліком є те, що вони не враховують фактор розділимості навчальної вибірки. Дана характеристика показує, наскільки сигнатури різних класів не перетинаються між собою. Низький рівень розділимості властивий змішаній навчальній вибірці. Таким чином, розділимість суттєво впливає на точність класифікації. Одним з варіантів підвищення розділимості є кластеризація навчальної вибірки. Отже, **метою** даного дослідження була розробка методики кластеризації навчальної вибірки, яка дозволяє підвищити точність класифікації земного покриву безпосередньо за рахунок підвищення розділимості навчальної вибірки. Таким чином **завдання** цього дослідження наступні: 1) розробити метод оцінювання розділимості навчальної вибірки; 2) розробити методику кластеризації навчальної вибірки; 3) перевірити ефективність розробленої методики, виконавши експериментальну класифікацію земних покривів із застосуванням розробленої методики. В якості експерименту було отримано по дві класифікації для кожної з двох обраних територій дослідження: одна класифікація до застосування методики, а друга – після. Було залучено шість класів в кожному з експериментів. Навчальну вибірку було відібрано для кожного з класів. **Висновки**. Після застосування розробленої методики було зафіксовано підвищення розділимості навчальної вибірки, яке зафіксовано розрбленим індексом розділимості. В свою чергу, це призвело до підвищення точності класифікації. Для першого експерименту це засвідчено підвищенням загальної точності класифікації та капа-коефіцієнта на 20% (з 63% до 83%) та 21% (з 60% до 81%) відповідно. А для другого експеримента підвищення загальної точності класифікації та капа-коефіцієнта становило 4% (з 77 до 81%) та 5% (з 66% до 71%) відповідно.

Ключові слова: класифікація; керована класифікація; некерована класифікація; кластеризація; ДЗЗ; навчальна вибірка; розділимість навчальної вибірки.

Андреєв Артем Андрійович – доктор філософії, молодший науковий співробітник, ДУ "Науковий центр аерокосмічних досліджень Землі ІГН НАН України", Київ, Україна.

Артюшин Леонід Михайлович – д-р техн. наук, проф., головний науковий співробітник, Державний науково-дослідний інститут авіації, Київ, Україна.

Artem Andreiev – PhD, Junior Researcher, Scientific Centre for Aerospace Research of the Earth, Kyiv, Ukraine,
e-mail: artem.a.andreev@gmail.com, ORCID: 0000-0002-6485-449X, Scopus Author ID: 57203280237.

Leonid Artiushyn – Doctor of Technical Sciences, Professor, State Research Institute of Aviation, Kyiv, Ukraine,
e-mail: artleonid2017@gmail.com, ORCID: 0000-0002-7488-7244, Scopus Author ID: 6602317317.